

## PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-283123

(43)Date of publication of application : 23.10.1998

(51)Int.Cl.

G06F 3/06

G06F 3/06

G06F 1/32

G06F 1/26

(21)Application number : 09-085364

(71)Applicant : XING:KK

BROTHER IND LTD

(22)Date of filing : 03.04.1997

(72)Inventor : MIYAKOSHI MITSUNARI

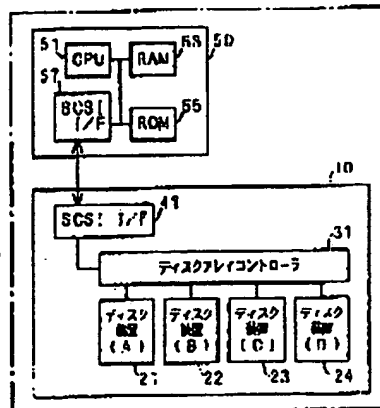
KUNO TAKUMA

## (54) STORAGE DEVICE ARRAY

## (57)Abstract:

**PROBLEM TO BE SOLVED:** To efficiently supply power and to suppress the degradation of a storage device and the consumption of the power by supplying the power to a standby device in the case of judging that an error is generated in one of minimum storage devices capable of generating the data of a read object.

**SOLUTION:** A disk array 10 is provided with disk devices 21-24 as the four storage devices to be parallelly operated and a disk array controller 31 capable of individually controlling the respective disk devices 21-24, etc. Then, in the disk array 10, control for supplying the power to the three disk devices 21-23 equivalent to the minimum storage devices capable of generating the data of the read object and not supplying the power to the disk device 24 equivalent to the standby device is normally performed. Then, in the case of a condition that the error is generated for one of the disk devices 21-23, the control is performed so as to supply the power to the disk device 24 for the first time.



## LEGAL STATUS .

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision]

BEST AVAILABLE COPY

of rejection]

[Date of requesting appeal against examiner's  
decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平10-283123

(43) 公開日 平成10年(1998)10月23日

(51) Int. Cl. <sup>6</sup>	識別記号	F I
G 0 6 F 3/06	3 0 5	G 0 6 F 3/06 3 0 5 C
	5 4 0	5 4 0
1/32		1/00 3 3 2 B
1/26		3 3 4 G

審査請求 未請求 請求項の数 4 O L (全 9 頁)

(21) 出願番号 特願平9-85364

(22) 出願日 平成9年(1997)4月3日

(71) 出願人 396004833

株式会社エクシング

名古屋市瑞穂区塩入町18番1号

(71) 出願人 000005267

ブラザー工業株式会社

愛知県名古屋市瑞穂区苗代町15番1号

(72) 発明者 宮腰 光成

愛知県名古屋市中区錦3丁目10番33号 株式会社エクシング内

(72) 発明者 久野 琢磨

愛知県名古屋市中区錦3丁目10番33号 株式会社エクシング内

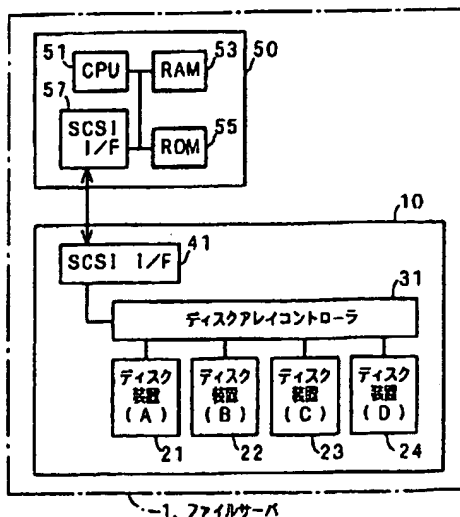
(74) 代理人 弁理士 足立 勉

(54) 【発明の名称】 記憶装置アレイ

(57) 【要約】

【課題】 例えば複数の記憶装置の中の所定台からデータが読み出せれば読出対象のデータを生成可能であり、残りは予備的な記憶装置として捉えることのできる RAIDにおいて、効率的な電源供給を行って、ディスク装置等の記憶装置の劣化及びその記憶装置による電力の消費を抑える。

【解決手段】 (A) ~ (C) のディスク装置 21 ~ 23 の3台からデータが読み出せれば読出対象のデータを生成可能な場合、それらにおいてエラーが発生せず正常にデータ読み出しができるのであれば予備的なディスク装置 (D) 24 については電源供給せず、エラーが発生して予備的なディスク装置 (D) 24 が必要となつて初めて電源供給する。この電源供給制御はスイッチング制御部 250 を介した CPU 201 からのポート制御により、各電源ライン 251、252、253、254 を個別に ON/OFF して実行する。



## 【特許請求の範囲】

【請求項1】 アレイコントローラによって複数台の記憶装置を並行してアクセス可能であり、当該複数台の記憶装置にデータを分散させて記憶し、当該複数台の記憶装置のうちの少なくとも1台の記憶装置には冗長データを記憶することによって、メインコンピュータからの指示によりデータ読出処理を行う際には、前記複数台の記憶装置のうちの所定数台の記憶装置からのデータに基づき読出対象のデータを生成可能な記憶装置アレイにおいて、

前記並行してアクセス可能な記憶装置について個別にステータスを取得可能なステータス取得手段と、

前記読出対象のデータを生成可能な最低限の記憶装置についてののみ電源供給を行い、それ以外の記憶装置は予備装置として電源供給を基本的に行われないように制御する電源供給制御手段とを備え、

さらに、当該電源供給制御手段は、前記ステータス取得手段によって取得したステータスに基づいて、前記読出対象のデータを生成可能な最低限の記憶装置のいずれかにおいてエラーが発生したと判断した場合には、前記予備装置への電源供給を行なうよう構成されており、

前記アレイコントローラは、前記エラーが発生した記憶装置の代わりに前記電源供給の開始された予備装置を読出対象データの生成用の記憶装置として扱うよう構成されていることを特徴とする記憶装置アレイ。

【請求項2】 前記電源供給制御手段が、前記取得したステータスに基づき前記読出対象のデータを生成可能な最低限の記憶装置のいずれかでエラーが発生したと判断した場合に前記予備装置への電源供給を行ない、前記アレイコントローラが、前記エラーが発生した記憶装置の代わりに前記電源供給の開始された予備装置を読出対象データの生成用の記憶装置として扱う一連の電源供給制御処理を所定時間毎に実行するよう構成されていることを特徴とする請求項1に記載の記憶装置アレイ。

【請求項3】 検査用データを記憶しており、当該検査用データを用いて前記記憶装置へのデータ書き込み及び読み出しを実行した場合の前記ステータス取得手段によって取得したステータスに基づき、前記エラー発生を判断を実行するよう構成されていることを特徴とする請求項2に記載の記憶装置アレイ。

【請求項4】 前記ステータス取得手段は、プロトコルコントローラであり、前記並行してアクセス可能な記憶装置については個別にプロトコルコントローラが設けられていることを特徴とする請求項1乃至3のいずれかに記載の記憶装置アレイ。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】 本発明は、アレイコントローラによって複数台の記憶装置を制御し、メインコンピュータからのアクセスに対しては1台の記憶装置に見せか

けて応答するよう構成された記憶装置アレイに関する。

## 【0002】

【従来の技術】 従来、高速アクセスを可能とし、かつ高信頼性を保証する記憶装置としてディスクアレイが提案されてきた。ディスクアレイは、例えばハードディスク装置等の小型のディスク装置を複数備えることで、大型ディスク装置に対抗して高速アクセスを可能にする方式として体系化されている。この体系化された方式はRAID (Redundant Arrays of Inexpensive Disks) として、David A Pattersonら3人によって発表され (U.C. Berkeley Report NO. UCB/CSD 87/39参照)、現在ではレベル1からレベル5までが基本的な構成として考えられている。ここでは、本発明に関係する代表的なものとして、レベル3のRAIDについて説明する。

【0003】 レベル3のRAIDでは、1台のパリティ用ディスク装置と2台以上のデータ用ディスク装置からなる複数台のディスク装置を並行にアクセスできるようになっている。そして、メインコンピュータからの要求によってデータを記憶するときには、そのデータを1バイト単位に分割し、データ用ディスク装置のそれぞれに分散して記憶すると共に、それらの分割データの排他的論理和であるパリティデータを計算してパリティ用ディスク装置に記憶する。つまり、データは、並行してアクセスできる複数台のディスク装置のそれぞれに分散して記憶されることになる。

【0004】 このように、本来1台のディスク装置に書き込まれるはずのデータが複数台のディスク装置に分散して書き込まれるため、トータルで考えた場合、1台のディスク装置に対する読み書きの頻度や読み出すデータ量が低減されるため、全体として見れば処理速度の向上が図られるのである。また、複数台のディスク装置のうちの1台をパリティ用ディスク装置とするため、1台のデータ用ディスク装置からのデータが読み出せない場合であっても、残りのデータ用ディスク装置からのデータとパリティ用ディスク装置からのパリティデータに基づいて、その1台のデータ用ディスク装置から読み出されるはずのデータを生成することができ、信頼性の向上が図られるのである。

【0005】 そして、このディスクアレイ装置においては、データ読み出しのパフォーマンスの低下を防止するために、備えている複数台のディスク装置を全て稼働させて読み出しがいつでも開始できるようにされていた。ディスク装置の稼働中は電源が供給され続けている。

## 【0006】

【発明が解決しようとする課題】 しかしながら、このように全てのディスク装置を稼働させていても、常に全てのディスク装置からデータ読み出さなければメインコンピュータから要求されたデータを生成できないわけではない。例えば、上述したレベル3のRAIDの場合でいえば全てのデータ用ディスク装置から正常にデータを

読みせればパリティデータがなくてもデータ生成ができるため、パリティ用ディスク装置からのデータ読み出しはしなくても構わなくなるのである。つまり、この場合のパリティ用ディスク装置は、予備的なディスク装置として捉えることができる。

【0007】したがって、結果的には最終的なデータ生成には不要なディスク装置に対しても電源供給を行っていることとなり、ディスク装置の劣化が激しくなり、また、無駄な電力が消費されるという問題が生じていた。本発明は、例えば複数の記憶装置の中の所定台からデータが読み出せれば読出対象のデータを生成可能であり、残りは予備的な記憶装置として捉えることのできるRAIDにおいて、効率的な電源供給を行って、ディスク装置等の記憶装置の劣化及びその記憶装置による電力の消費を抑えることを目的とする。

【0008】

【課題を解決するための手段及び発明の効果】上記目的を達成するためになされた請求項1に記載の記憶装置アレイは、アレイコントローラによって複数台の記憶装置を並行してアクセス可能であり、当該複数台の記憶装置にデータを分散させて記憶し、当該複数台の記憶装置のうちの少なくとも1台の記憶装置には冗長データを記憶することによって、メインコンピュータからの指示によりデータ読出処理を行う際には、前記複数台の記憶装置のうちの所定数台の記憶装置からのデータに基づき読出対象のデータを生成可能な記憶装置アレイにおいて、前記並行してアクセス可能な記憶装置について個別にステータスを取得可能なステータス取得手段と、前記読出対象のデータを生成可能な最低限の記憶装置についてのみ電源供給を行い、それ以外の記憶装置は予備装置として電源供給を基本的に行われないように制御する電源供給制御手段とを備え、さらに、当該電源供給制御手段は、前記ステータス取得手段によって取得したステータスに基づいて、前記読出対象のデータを生成可能な最低限の記憶装置のいずれかにおいてエラーが発生したと判断した場合には、前記予備装置への電源供給を行なうよう構成されており、前記アレイコントローラは、前記エラーが発生した記憶装置の代わりに前記電源供給の開始された予備装置を読出対象データの生成用の記憶装置として扱うよう構成されていることを特徴とする。

【0009】本発明の記憶装置アレイでは、アレイコントローラによって複数台の記憶装置を並行してアクセス可能である。なお、 $n$  ( $n$ は2以上の自然数、以下、同じ。) 台の記憶装置が並行してアクセス可能である場合、並行してアクセス可能な $n$ 系統に1台ずつの記憶装置が属し、 $n$ 台の記憶装置が並行してアクセス可能となっていることも考えられるし、並行してアクセス可能な $n$ 系統に2台以上の記憶装置が属し、その各系統から1台ずつを選択した $n$ 台の記憶装置が並行してアクセス可能となっていることも考えられる。

【0010】アレイコントローラによって、データは複数台の記憶装置のそれぞれに分散して記憶され、その複数台の記憶装置のうちの少なくとも1台の記憶装置にはパリティデータ等の冗長データが記憶される。冗長データを利用することによって、メインコンピュータからの指示によるデータ読出処理では、複数台の記憶装置のうちの所定数台の記憶装置からのデータに基づき読み出し対象のデータを生成することが可能である。例えばパリティデータが冗長データとして用いられる場合、所定数は $(n-1)$ となり、 $(n-1)$ 台の記憶装置からのデータに基づき読出対象のデータを生成可能である。

【0011】そして、本記憶装置アレイでは、電源供給制御手段が、読出対象のデータを生成可能な最低限の記憶装置についてのみ電源供給を行い、それ以外の記憶装置は予備装置として電源供給を基本的に行われないように制御するのであるが、並行してアクセス可能な記憶装置についてステータス取得手段が個別にステータスを取得可能であり、電源供給制御手段は、その取得したステータスに基づいて、読出対象のデータを生成可能な最低限の記憶装置のいずれかにおいてエラーが発生したと判断した場合には、予備装置への電源供給を行なう。そして、アレイコントローラは、エラーが発生した記憶装置の代わりに電源供給の開始された予備装置を読出対象データの生成用の記憶装置として扱う。

【0012】このように、複数の記憶装置の中の所定台からデータが読み出せれば読出対象のデータを生成可能な場合、それらにおいてエラーが発生せず正常にデータ読み出しができるのであれば予備的な記憶装置については電源供給せず、エラーが発生して予備的な記憶装置が必要となって初めて電源供給することとなるので、効率的な電源供給を行える。したがって、データ読み出しに用いない場合の予備的な記憶装置の劣化及びその記憶装置による電力の消費を抑えることができる。

【0013】なお、上述したように、本発明の記憶装置アレイでは、電源供給制御手段が、取得したステータスに基づき読出対象のデータを生成可能な最低限の記憶装置のいずれかでエラーが発生したと判断した場合に予備装置への電源供給を行ない、アレイコントローラが、エラーが発生した記憶装置の代わりに電源供給の開始された予備装置を読出対象データの生成用の記憶装置として扱うという一連の「電源供給制御処理」を実行するのであるが、この処理は、例えば実際にメインコンピュータからの指示によりデータ読出処理を行う際に実行してもよいが、メインコンピュータからの指示によらず、記憶装置アレイ自身が所定時間毎に実行してもよい。

【0014】つまり、メインコンピュータからの指示によりデータ読出処理を行う際に実行した場合には、エラーが生じて初めて予備装置への電源供給を開始し、その予備装置からデータを読み出して、読出対象のデータを生成することとなるため、レスポンスはどうしても低下

してしまう。したがって、そのようなレスポンスの低下の考慮が不要な時期に同様の制御処理を実行しておけばよい。そうすれば、前もってエラーが生じた記憶装置の代わりに予備装置をデータ読み出しの対象として扱う準備が完了するため、実際にメインコンピュータからの指示によりデータ読出処理を行う際のレスポンス低下を防止することができる。

【0015】なお、このように、メインコンピュータからの指示によらず、記憶装置アレ自身所定時間毎に電源供給制御処理を実行する場合には、例えば、検査用データを記憶しており、当該検査用データを用いて記憶装置へのデータ書き込み及び読み出しを実行した場合のステータス取得手段によって取得したステータスに基づき、エラー発生を判断を実行するよう構成することが考えられる。

【0016】また、ステータス取得手段は、例えばプロトコルコントローラを用い、並行してアクセス可能な記憶装置については個別にプロトコルコントローラを設けるように構成することが考えられる。例えば一般的なRAIDにおいては、記憶装置としてハードディスク装置が用いられ、アレコントローラとハードディスク装置とをSCSIバスで接続しているため、SCSIプロトコルコントローラ（SPC）が採用されることが多い。

【0017】

【発明の実施の形態】以下、本発明を具体化した一実施形態を図面を参照して説明する。図1は、本発明の記憶装置アレをファイルサーバ1に適用した概略構成を示すブロック図である。

【0018】図1に示すように、ファイルサーバ1は、「記憶装置アレ」としてのディスクアレ10と、メインコンピュータ50とから構成されている。メインコンピュータ50は、制御手段としてのCPU51と、「記憶手段」としてのRAM53と、プログラム記憶手段としてのROM55と、ディスクアレ10と接続するためのSCSIインターフェース57とを備えている。

【0019】一方、ディスクアレ10は、並列に動作する4台の「記憶装置」としてのディスク装置21、22、23、24と、それら各ディスク装置21～24を個別に制御可能な「アレコントローラ」としてのディスクアレコントローラ31と、メインコンピュータ50と接続するためのSCSIインターフェース41とを備えている。このように、並列に動作する4台のディスク装置21～24を備えるディスクアレ10を系統数4のディスクアレ10という。

【0020】なお、以下の説明では、4台のディスク装置21～24を区別するために、ディスク装置(A)21、ディスク装置(B)22、ディスク装置(C)23及びディスク装置(D)24と記載することにする。また、ディスク装置21～24は、いわゆる物理的なハー

ドディスクドライブとそれを制御するコントロールボードとが一体化されたものである。

【0021】続いて、ディスクアレコントローラ31の内部構成について図2を参照して説明する。なお、図2は、4台のディスク装置21～24に対する制御関連の構成だけを示したものであり、それ以外の構成は省略してある。ディスクアレコントローラ31は、本コントローラ全体の制御を司る制御手段としてのCPU201と、各種処理データを格納しておくRAM200と、CPU201が実行する動作プログラムなどを格納しておくROM202と、4本のSCSIラインを構成するSCSIバス(A)101、SCSIバス(B)102、SCSIバス(C)103及びSCSIバス(D)104に対してのリード及びライトデータをそれぞれ格納しておくための4つのバッファであるバッファ(A)211、バッファ(B)221、バッファ(C)231及びバッファ(D)241と、各SCSIラインとの間でのコマンド、データ及びステータスの送受信を行う4つのSCSIプロトコルコントローラ（SPC）であるSPC(A)213、SPC(B)223、SPC(C)233及びSPC(D)243とを備えている。

【0022】さらに、ディスクアレコントローラ31は、各SCSIラインからのデータリード及び各SCSIラインへのデータライト時にデータをDMA（ダイレクト・メモリ・アクセス）転送するためのDMAコントローラを各ライン毎に備えている。すなわち、DMAC(A)212、DMAC(B)222、DMAC(C)231及びDMAC(D)241である。

【0023】また、各SCSIラインには対応するディスク装置が接続されている。すなわち、SCSIバス(A)101にはディスク装置(A)21、SCSIバス(B)102にはディスク装置(B)22、SCSIバス(C)103にはディスク装置(C)23、そしてSCSIバス(D)104にはディスク装置(D)24が接続されている。

【0024】また、4本のSCSIラインであるSCSIバス(A)101、SCSIバス(B)102、SCSIバス(C)103及びSCSIバス(D)104に対応する電源ライン251、252、253、254は、CPU201のポートに割り振られており、スイッチング制御部250を介してCPU201からのポート制御により、各電源ライン251、252、253、254を個別にON/OFFすることができるようにされている。

【0025】ここで、最初に本実施形態のファイルサーバ1の動作の前提となる基本的な機能を説明する。本実施形態のファイルサーバ1においては、メインコンピュータ50から転送されたデータをディスクアレコントローラ31が1バイト単位に分割し、この場合は、ディスク装置(A)21、ディスク装置(B)22、ディス

ク装置 (C) 23の3台の同一セクタ上に順次書き込むと共に、それらのデータに対応するパリティを生成してディスク装置 (D) 24に書き込むよう構成されている。

【0026】従って、読出対象となるデータは、(A)～(C)の3台のディスク装置21～23からのデータを結合することにより生成できる。このとき、(A)～(C)の3台のディスク装置21～23のうちの1台にリードエラーが発生しても、残りの2台のディスク装置とディスク装置 (D) 24からのパリティデータに基づいて、リードエラーとなったディスク装置から読み出されるはずのデータを生成することが可能であり、読出対象となるデータを生成することができる。ファイルサーバ1のこのような基本的な機能を考えると、(A)～(C)のディスク装置21～23の3台が「読出対象のデータを生成可能な最低限の記憶装置」に相当し、ディスク装置 (D) 24が「予備装置」に相当することとなり、本実施形態のディスクアレイ10はレベル3のRAIDに相当する。

【0027】したがって、本実施形態のディスクアレイ10では、読出対象のデータを生成可能な最低限の記憶装置に相当する(A)～(C)のディスク装置21～23の3台については電源供給し、「予備装置」に相当するディスク装置 (D) 24については電源供給を行わないよう制御を通常は行っている。そして、(A)～(C)のディスク装置21～23の内のいずれか1台についてエラーが生じる状況となった場合に初めてディスク装置 (D) 24について電源供給を行うように制御するのである。これらの電源供給の実際の制御はCPU201が行なうのであるが、上述したように、スイッチング制御部250を介したCPU201からのポート制御により、各電源ライン251、252、253、254を個別にON/OFFしてこの電源供給制御を実行することとなる。

【0028】図3には、この電源供給、停止にかかる判断及び指示内容を示すフローチャートである。本処理は、メインコンピュータ50からのデータ読み出し要求などがない状態において所定時間毎に実行される。この処理ルーチンは、現在使用しているSCSIラインに対応するSPCからのステータスをチェックし、その結果に応じて電源供給の切替などを制御するルーチンである。ここでは、上述したように(A)～(C)のディスク装置21～23の3台については電源供給し、ディスク装置 (D) 24については電源供給を行わないような制御を基本的に実行していると前提とする。したがって、SCSIバス (A) 101、SCSIバス (B) 102、SCSIバス (C) 103の3本のバスにはそれぞれ対応する電源ライン251、252、253がONに制御されていて電源供給がされており、SCSIバス (D) 104に対応する電源ライン254はOFFに制

御されていて電源供給がされていない状況である。したがって、この場合には、SPC (A) 213、SPC (B) 223及びSPC (C) 233からのステータスをチェックすることとなる。

【0029】最初のステップS1では、CPU201はSPC (A) 213から入力したステータスをチェックしてエラー発生がどうかを判断する。なお、この場合には、例えばRAM200に所定の検査用データを記憶しており、その検査用データを用いてディスク装置 (A) 21へのデータ書き込み及び読み出しを実行し、その際のSPC (A) 213から入力したステータスに基づいてエラー発生を判断を実行する。なお、この処理に際しては、RAM200内の検査用データはバッファ (A) 211に送られ、DMAC (A) 212によってSPC (A) 213に転送されることとなる。

【0030】そして、エラー発生の場合には (S1: YES)、S4へ移行して、対応するSCSIバス (A) 101への電源供給をOFFするようにスイッチング制御部250に指示する。続くS7では、予備装置であるディスク装置 (D) 24が接続されているSCSIバス (D) 104への電源供給をONするようにスイッチング制御部250に指示する。そして、S8では、上記S7で電源供給を開始したSCSIバス (D) に接続されているディスク装置 (D) 24を、エラーが発生したSCSIバス (A) 101に接続されていたディスク装置 (A) 21の代わりに、読出対象データの生成用のディスク装置として扱うように設定してから、本処理ルーチンを終了する。

【0031】一方、SPC (A) 213から入力したステータスに対するチェックではエラー発生がなかった場合には (S1: NO)、S2へ移行して、今度はSPC (B) 223から入力したステータスに対するチェックを行い、上述したSPC (A) 213の場合と同様の処理を行なう。すなわち、エラー発生の場合には (S2: YES)、S5へ移行して、対応するSCSIバス (B) 102への電源供給をOFFしてからS7へ移行する。S7以降の処理は上述したので詳しい説明は省略するが、S8では、この場合、ディスク装置 (D) 24を、エラーが発生したSCSIバス (B) 102に接続されていたディスク装置 (B) 22の代わりに、読出対象データの生成用のディスク装置として扱うこととなる。

【0032】また、SPC (B) 223から入力したステータスに対するチェックではエラー発生がなかった場合には (S2: NO)、S3へ移行して、今度はSPC (C) 233から入力したステータスに対するチェックを行い、エラー発生の場合には (S3: YES)、S6へ移行して、対応するSCSIバス (C) 103への電源供給をOFFしてからS7、S8の処理を実行する。この場合、S8では、ディスク装置 (D) 24を、エラ

ーが発生したSCSIバス(C)103に接続されていたディスク装置(C)23の代わりに、読出対象データの生成用のディスク装置として扱うこととなる。

【0033】もちろん、SPC(C)233から入力したステータスに対するチェックでもエラー発生がなかった場合には(S3:NO)、そのまま本処理ルーチンを終了する。なお、CPU201が「電源供給制御手段」に相当し、上述の図3のフローチャートにて示した処理が電源供給制御手段としての処理に相当する。

【0034】このように、本実施形態のディスクアレイ10においては、(A)～(C)のディスク装置21～23の3台からデータが読み出せば読出対象のデータを生成可能な場合、それらにおいてエラーが発生せず正常にデータ読み出しができるのであれば予備的なディスク装置(D)24については電源供給せず、エラーが発生して予備的なディスク装置(D)24が必要となって初めて電源供給することとなるので、効率的な電源供給を行える。したがって、データ読み出しに用いない場合の予備的なディスク装置(D)24の劣化及びそのディスク装置(D)24及び接続されているSCSIバス(D)104による電力消費を抑えることができる。

【0035】以上、本発明はこのような実施形態に何等限定されるものではなく、本発明の主旨を逸脱しない範囲において種々なる形態で実施し得る。例えば、上記実施形態のディスクアレイ10においては、メインコンピュータ50からの指示によらず、ディスクアレイ10自身が所定時間毎に電源供給制御を実行するようにしたが、実際にメインコンピュータ50からの指示によりデータ読出処理を行う際に実行してもよい。つまり、上述した実施形態では、RAM200内の検査用データを用いてディスク装置へのデータ書き込み及び読み出しを実行し、その際のSPCから入力したステータスに基づいてエラー発生を判断を実行したが、メインコンピュータ50から指示されたデータを読み出し、その際のSPCから入力したステータスに基づいてエラー発生を判断を実行すればよい。

【0036】但し、メインコンピュータ50からの指示によりデータ読出処理を行う際に実行した場合には、エラーが生じて初めて予備装置(上記実施形態ではディスク装置(D)24)の接続されているSCSIバス(D)104への電源供給を開始し、そのディスク装置(D)24からデータを読み出して、読出対象のデータを生成することとなるため、レスポンスはどうしても低下してしまう。したがって、上記実施形態のようにすれば、レスポンスの低下の考慮が不要な時期に同様の制御処理を実行することができる。つまり、前もってエラーが生じたディスク装置の代わりに予備装置であるディスク装置(D)24をデータ読み出しの対象として扱う準備が完了するため、実際にメインコンピュータ50からの指示によりデータ読出処理を行う際のレスポンス低下

を防止することができるのである。

【0037】なお、上記実施形態では、4本のSCSIバス101～104に対してそれぞれ1台ずつのディスク装置21～24が接続される構成であったが、各SCSIバス101～104に対してそれぞれ2台以上のディスク装置を接続するような構成であってもよい。

【0038】また、上記実施形態では、レベル3のRAIDについて説明したが、その他のレベルのRAIDであっても、「読出対象のデータを生成可能な最低限の記憶装置」以外に「予備装置」として捉えることができる装置を備える構成の場合には同様に適用できる。

【0039】ここでは、上述したレベル3も含めてディスクアレイのその他のレベルについて簡単に説明しておく。以下の説明では、ディスクを何台並列に並べるかという数を、パラレル数と呼び、 $p$ という変数で表すこととする。但し、パラレル数 $p$ には、冗長データを格納するためのディスクは含まない。

【0040】なお、図4(A)にはレベル0のRAID、(B)にはレベル1のRAIDの原理説明図を示し、図5(A)にはレベル2のRAID、(B)にはレベル3のRAIDの原理説明図を示している。また、図6(A)にはレベル4のRAID、(B)にはレベル5のRAIDの原理説明図を示している。

【0041】RAIDレベル0は、単にディスクをパラレル動作させ、データを分散記憶させるものである。信頼性向上の効果はなく、高速化の効果しかない。分散の単位は、ビット単位でもバイト単位でもセクタ単位でも良く、特に限定されない。RAIDレベル0は正確にはRAIDではないが、対比のためによく例に挙げられる。

【0042】RAIDレベル1は、ミラーリングとも呼ばれ、2つのディスクに同一のデータを書き込み、読み出すときはどちらか一方のディスクから読み出す。片方のディスクが壊れても、データは失われない。RAIDレベル1は高速化の効果はないが信頼性が増大する。

【0043】RAIDレベル2は、ハミングコードなどの冗長符号(誤り訂正符号)を用いるもので、レベル1が通常のディスクと比べて2倍のディスクを必要とするのに対し、レベル2は2倍までは要らない。データはビット単位あるいはバイト単位で分散させ、冗長符号と合わせて記録する。冗長符号の選び方で色々な方法が考えられるが特に冗長符号として単純なパリティを用いたものが次のレベル3である。他の冗長符号を用いたものはあまり実用価値がないので、レベル2が使用されることは希である。

【0044】RAIDレベル3は、バイト単位でデータを分散させ、パリティを付加してディスクに格納するので、高速化、信頼性向上の両方の効果がある。反面、ディスクのセクタサイズ $\times p$ 個のデータが集まらないと読み書きができないので、ディスクを単体で使う場合に



比べて、アクセス単位が大きくなる。データを小さな単位でアクセスするような用途には不向きである。

【0045】RAIDレベル4は、この点を改善したもので、ディスクのセクタ単位でデータを分散させるものである。データはセクタサイズ単位に分割して、各ディスクに順に格納するのだが、p個のセクタごとにパリティデータを計算して冗長データディスクに格納する。RAIDレベル3ではセクタ単位×pのデータが最小アクセス単位だったが、RAIDレベル4ではセクタ単位で読み書きが可能である。しかしセクタを書き換えるときは元のパリティデータと元のディスクデータを読み出して、新パリティを計算し直し、これをパリティとして書き戻す必要があるため、通常より余分な動作が必要となる。またパリティを格納しているディスクにアクセスが集中するため、ここがボトルネックになるという欠点もある。

【0046】RAIDレベル5はこの点を改善したもので、パリティを格納するディスクを、ブロックごとに回転させることで、特定のディスクにアクセスが集中するのを防止している。さらには、RAID以外のものであっても、「読出対象のデータを生成可能な最低限の記憶装置」以外に「予備装置」として捉えることができる装置を備える構成の場合には同様に適用できる。

【図面の簡単な説明】

【図1】 本発明の記憶装置アレイをファイルサーバに適用した概略構成を示すブロック図である。

【図2】 実施形態のディスクアレイコントローラの内

部構成を示すブロック図である。

【図3】 実施形態のディスクアレイコントローラが実行する電源供給制御処理を示すフローチャートである。

【図4】 RAIDレベル0及びRAIDレベル1の原理説明図である。

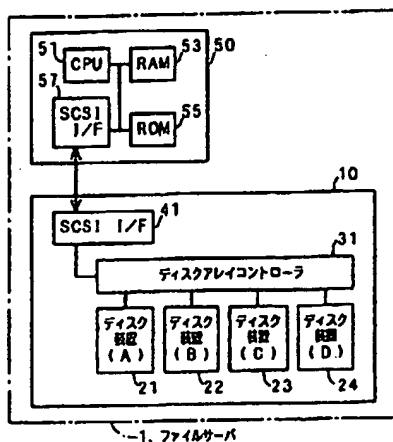
【図5】 RAIDレベル2及びRAIDレベル3の原理説明図である。

【図6】 RAIDレベル4及びRAIDレベル5の原理説明図である。

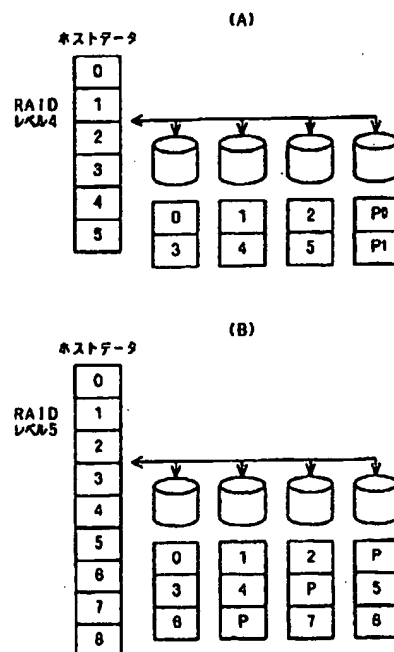
【符号の説明】

1…ファイルサーバ                      10…ディスクアレイ  
21～24…ディスク装置                  31…ディスクアレイコントローラ  
41…SCSIインターフェース              50…メインコンピュータ  
51…CPU                                  53…RAM  
55…ROM                                  57…SCSIインターフェース  
101～104…SCSIバス                  200…ROM  
201…CPU                                  202…RAM  
211, 221, 231, 241…バッファ  
212, 222, 232, 242…DMAコントローラ  
213, 223, 233, 243…SCSIプロトコルコントローラ (SPC)  
250…スイッチング制御部  
251～254…電源ライン

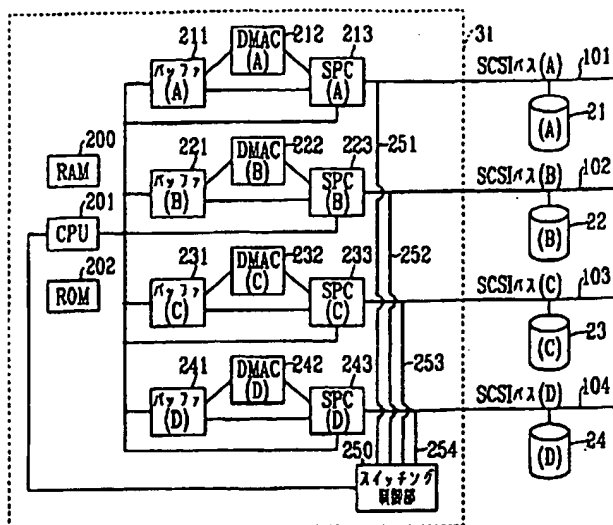
【図1】



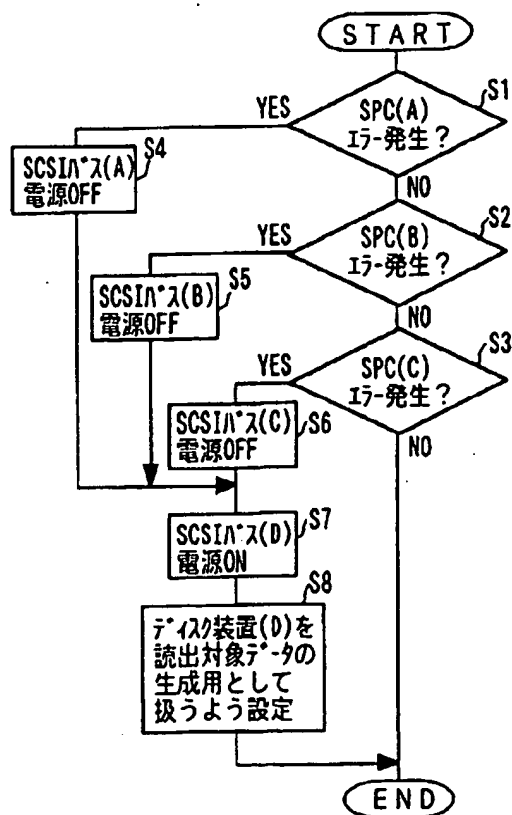
【図6】



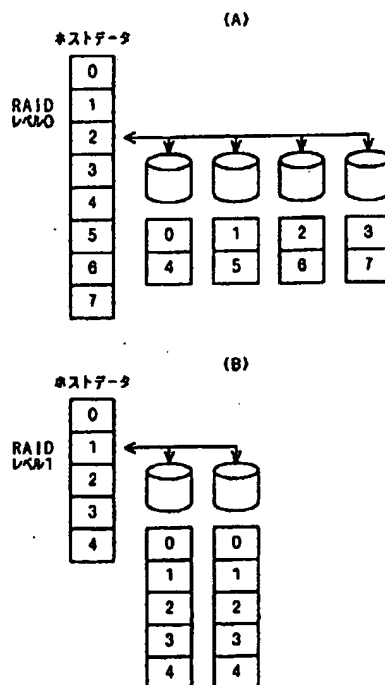
【図2】



【図3】



【図4】



〔図5〕

